Pers Ubiquit Comput (2014) 18:385–396
DOI 10.1007/s00779-013-0657-z

ORIGINAL ARTICLE

Routing with multi-level cross-community social groups in mobile opportunistic networks

Fan Li · Lunan Zhao · Chao Zhang · Zhenmin Gao · Yu Wang

Received: 20 August 2012 / Accepted: 15 January 2013 / Published online: 11 April 2013

Abstract Mobile opportunistic networks (MONs) are intermittently connected networks, such as pocket switched networks formed by human-carried mobile devices. Routing in MONs is very challenging as it must handle network partitioning, long delays, and dynamic topology. Flooding is a possible solution but with high costs. Most existing routing methods for MONs avoid the costly flooding by selecting one or multiple relays to deliver data during each encounter. How to pick the "good" relay from all encounters is a non-trivial task. To achieve efficient delivery of messages at low costs, in this paper, we propose a novel group-based routing protocol in which the relay node is selected based on multi-level cross-community social group information. We apply a simple group formation method to both historical encounters (social relationships in physical world) and/or social profiles of mobile users (social relationships in social world) and build multi-level cross-community social groups, which summarize the wide range of social relationships among all mobile participants. Our simulations over several real-life data sets demonstrate the efficiency and effectiveness of the proposed method by comparing it with several existing MON routing schemes.

F. Li (✉) · L. Zhao · C. Zhang · Z. Gao
School of Computer Science, Beijing Institute of Technology, Beijing 100081, China
e-mail: fli@bit.edu.cn

Y. Wang (✉)
Department of Computer Science, University of North Carolina at Charlotte, Charlotte, NC 28223, USA
e-mail: yu.wang@uncc.edu

Keywords Routing · Relay selection · Social group · Cross-community · Multi-level · Mobile opportunistic networks

1 Introduction

Mobile opportunistic network (MON) is one of the emerging communication paradigms in wireless mobile communications. MONs are commonly defined as a type of mobile networks where communication is challenged by sporadic and intermittent contacts as well as frequent disconnections and reconnections, and where the assumption of the existence of an end-to-end path between the source and the destination is relinquished. Examples include pocket switched networks (PSNs) [1] or mobile social networks (MSNs) [2], which are comprised of human-carried mobile devices moving in a restricted physical space and use occasional contact opportunities to deliver data. Intermittent connectivity in MONs results in lack of instantaneous end-to-end paths, large transmission delays, and unstable network topology. These characteristics make the classical mobile ad hoc routing protocols not applicable for MONs; therefore, many opportunity-based routing protocols [3–10] are proposed recently for MONs or general delay tolerant networks.

Most of existing opportunity-based routing methods for MONs share the same principle, "store and forward", to handle intermittent connectivity. If there is no connection available at a particular time, the current node can store and carry the data until it encounters other nodes. When the node has such a forwarding opportunity, all encountered nodes could be the candidates to relay the data. Therefore, relaying selection and forwarding decision need to be made by the current node based on certain forwarding strategy.

The simplest routing method is *epidemic routing* [3], in which a node forwards copies of message to any nodes it encounters. This flooding-based method can guarantee the best delivery ratio, but with possibly huge message overheads. To reduce the overheads, many routing methods restrict the number of message replicas in the network to a certain constant (such as in *Spray and Wait* [5]) or just one (such as in *SimBet* [7]) or a small one by only replicating the message when certain condition is met (such as in *delegation forwarding* [6]). We call the method which allows multiple replicas and the method which allows a single replica as multi-copy routing and single-copy routing, respectively.

Forwarding decision (or replicating decision) and relay selection in these routing protocols usually rely on comparisons between per-node metrics. For example, in *FRESH* [8], the current node forwards if it encounters another node which has met the destination more recently than it does, and if multiple nodes satisfy such a condition during encounter, it just selects the one which has met the destination most recently as the relay; in *Greedy-Total* [9], the node forwards if it meets nodes with a higher contact frequency, it picks the one with highest frequency as the relay. In addition to these metrics which aim to estimate the delivery probability or expected delay to the destination node, there are also certain social metrics (such as community and centrality) which can be used to assist forwarding decision and relay selection in recent *social-based* approaches [4, 7, 11–13]. For example, in *SimBet* [7], the current node forwards if it encounters a node with higher social centrality and has more common neighbors with the destination; in *Bubble Rap* [4] the current node forwards data to the node with higher centrality following a hierarchical community structure. These social-based methods take the advantages of social relationships among nodes to make smarter forwarding decisions.

In this paper, we propose a new group-based routing protocol for mobile opportunistic networks, in which the relay node is selected based on social group information obtained from historical encounters or social profiles of mobile users. We introduce a simple but efficient formation method to build multi-level cross-community social groups, which summarizes the wide range of social relationships among all mobile participants. Notice that social relations and behaviors among mobile users are usually long-term characteristics and less volatile than node mobility. Our group-based routing method forwards the packet greedily toward the destination's social groups. We conduct extensive simulations using real-life tracing data [15–17] to compare the proposed method with several existing methods. Our simulation results demonstrate the efficiency and effectiveness of the proposed method.

This paper is organized as follows. Section 2 provides a brief review of existing opportunity-based routing protocols. Section 3 introduces the proposed routing method based on multi-level cross-community social graphs in details. Section 4 presents simulation results over different real-life data sets, and Sect. 5 concludes the paper. Preliminary results of this paper were appeared in [18, 19].

2 Related works

Mobile opportunistic networks are special cases of *delay/disruption tolerant networks* (DTNs) [20]. The major difference of DTNs from MONs is that mobility is often predictable or the future contact information is known. Irregularity of mobility pattern in MONs poses great challenges in the design of routing protocols. Here, we mainly focus on *opportunity-based routing* where messages are forwarded using available communication opportunities when nodes meet at the same place. By taking the advantages of mobility of intermediate nodes, it is expected to deliver the messages eventually, but with no guarantees.

Epidemic routing [3] floods copies of message to any nodes it encounters, thus can guarantee the delivery. However, it suffers from huge message overheads. Spyropoulos et al. [5] then proposed *Spray and Wait* routing which limits the total number of replicas of a message in the network to a constant x. The source of the message initially creates x replicas of the message. If a node u has $k > 1$ replicas and meets a node v with no replicas, u forwards half of its replicas to v and keeps the other half. Erramilli et al. [6] also proposed another way to reduce the total number of replicas, called *delegation forwarding*, in which the current node only forwards a replica to encountering nodes with highest-quality metric so far. In other words, a node will forward a message only if it encounters another node whose quality metric is greater than any nodes the message has yet met. Here, the quality metric can be defined in different ways as we will discuss it in the next paragraph for single-copy routing. All these three methods allow multiple replicas propagated in the network which can clearly improve the chance of delivery.

There are also many forwarding schemes which only allow one single copy of each message in the network, i.e., after forwarding the message to a single selected encountered node the current node will not forward anymore. Forwarding decision usually relies on certain type of quality metric, and the message is only forwarded to a node with higher quality metric. If during an encounter, there are multiple nodes with higher quality metric, only the one with highest-quality metric is selected as the relay. Examples include *FRESH* [8] (picking the node which has met the destination more recently), *Greedy-Total* [9] (picking the node with a higher encounter frequency to all other nodes), or *MobySpace* [21] (picking the node which has more location similarity with the destination).

Mobile devices in MONs are used and carried by people, whose behaviors are better described by social models. This opens the new possibilities of social-based routing [13] for MONs, in which the knowledge of social characteristics is used for making better forwarding decisions. For example, nodes with higher social centrality (more popular) are selected as relay nodes (such as in *SimBet* [7], *Bubble Rap* [4], and *friendship-based* routing [12]); or nodes within the same community (or social group) with the destination are preferred as relay nodes (such as in *Label* routing [11], *Bubble Rap* [4], and *friendship-based* routing [12]). Our proposed group-based method belongs to this category and uses the concept of social groups to extract underlying social relationships among all nodes. However, it is different from the community-based method (such as *Bubble Rap* [4]) by using a much simpler social group formation method. In addition, our group-based method supports exploring multi-level and cross-community social groups, which summarizes the wide range of social relationships among all mobile participants in both physical world (via historical encounters) and virtual social world (via social features from user profiles). This type of cross-community social graphs has not been well studied before. The only similar idea we could find is a hybrid social networking infrastructure proposed recently by Guo et al. [14], in which both online and opportunistic communities are considered and interlinked for the purposes of content sharing and information dissemination.

3 Multi-level cross-community social group-based forwarding

This section introduces our multi-level cross-community group-based routing protocol for mobile opportunistic networks in details. We start with a simple multi-level social group formation based on historical encounters and then follow with two versions of the proposed social group-based routing. Finally, we explain how to apply them with cross-community social groups where heterogeneous social groups are obtained from different sources (beyond historical encounters). Here, we assume that a set of n mobile nodes $V = \{v_1, v_2, \ldots, v_n\}$ and each possible data forwarding happens when two mobile nodes are in contact (*i.e.*, move within transmission range of each other). By recording contacts seen in the past, a *contact graph* G can be generated where each vertex denotes a mobile node (device or person who carries the device) and each edge represents one or more past meetings between two nodes. An edge in this contact graph conveys the information that two nodes encountered each other in the past. Such existence of an edge intends to have predictive capacity for future contacts. The contact graph can be constructed to record the encounters in a specific period of time by recording the time, the frequency and the duration of all encounters.

3.1 Simple multi-level social group formation

Since wireless devices are usually carried by people, it is natural to explore the social interactions among wireless devices in mobile opportunistic networks to design better forwarding strategy. *Social group* (or community) is an important concept from sociology [22–24]. A social group is usually defined as a group of interacting people living in a common location. Sociologists have studied the interactions between people in groups on many spatial and temporal scales [22–26] and shown that a member of a given social group is more likely to interact with another member of the same group than with a randomly chosen member of the population [25]. Therefore, social groups naturally reflect social relationships among people in social networks and implicitly define encounter patterns among wireless devices in mobile opportunistic networks. Our proposed group-based routing protocol hence uses social group information obtained from historical encountered data (a contact graph) to make its forwarding decision.

Constructing social groups from the encountered data could be done by using different methods. For example, there are several community detection algorithms [27–30] available for identifying social communities from the underlying contact graphs. Guo et al. [31] recently also proposed a group-mining method for extracting and formulating social groups from mobile social activity logging repository. However, most of these community detection or group-mining methods are relatively complex and not suitable for real-time executions in MONs. Instead, in this paper, we adopt a very simple social group formation method based on the number of past encounters among nodes. For any two nodes v_i and v_j, if there is more than t encounters between v_i and v_j in the past, they will be placed into the same group. Here, t is an adjustable threshold which defines how strong the social tie between two members is. In other words, given the contact graph G in the past, we only keep an edge between v_i and v_j when the number of their encounters is larger than or equal to t. For the graph G_t formed by all remaining edges, we treat each connected component as one social group. If two nodes are within the same group, there must be a path connecting them in G with all “strong” contact history. Another way to form the social group is letting each completed subgraph of G_t be a social group. This requires a

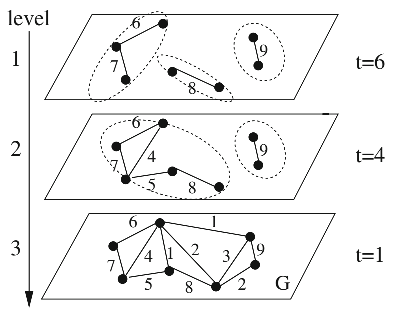

Fig. 1 *Multi-level social groups* formed from different values of t. The number labeled on each link is the number of past encounters between two endpoints. When $t = 1$, G_t is the original contact graph G

much stronger tie among group members, that is , any two members must directly have a "strong" contact history. In our simulations, we use the first method to define social groups. It is straightforward to compute the connected components of the contact graph G_t in linear time (in terms of the numbers of the nodes and edges of the graph) using either breadth-first search or depth-first search. By defining different values of t, we can construct multi-level social groups. Larger t leads to smaller groups with stronger ties. See Fig. 1 for illustration.

Note that our group-based forwarding algorithm does not rely on particular social group formation method. Social group information from other group/community detection algorithms or specified by users could be adopted in our model.

3.2 Routing with single-level social group

Having the knowledge of social groups could help routing protocol to choose better forwarding relays for particular destinations and hence improve the chance of delivery. Since it is believed that devices within the same social group have higher chances to encounter with each other, our group-based forwarding method intends to choose the member of social group of the destination as the preferred relay node.

Our group-based forwarding method (denoted by Group hereafter) works as follows. The current node v_i with a message M destined to v_d meets a set of nodes which means that they are capable to exchange messages. Assume that R is the set of nodes among them which do not hold message M, that is , all possible relay nodes at this particular time. If the destination v_d of message M is in R, v_i simply delivers the message to v_d. Otherwise, v_i looks for nodes within the same group of v_d (we use $g(v_d)$ to denote the member set of social group of v_d) as possible relay nodes. If there exists multiple such nodes, our method picks the one which has met the destination most recently as the relay node and forwards M to it. This is similar to the idea of *FRESH* [8]. If there is no any member of $g(v_d)$ in R, v_i continues holding M. Algorithm 1 shows the detailed description. In Group, social group information is used to increase the chance of meeting the destination, while the FRESH tries to deliver the message to destination as soon as possible.

Algorithm 1 Group-based Forwarding (Group)

Node v_i with a message M destined to v_d meets a set of nodes R which do not hold M.

```
if v_d ∈ R then
    Forward M to v_d
else
    if there exists v_j ∈ R within the same group with v_d,
    i.e., ∃v_j ∈ R ∩ g(v_d) then
        Let v_k be the node in R ∩ g(v_d) which has contacted
        v_d most recently
        Forward M to v_k
    end if
end if
```

Algorithm 2 Multi-level-Group-based Forwarding (mGroup)

Node v_i with a message M destined to v_d meets a set of nodes R which do not hold M. Information of m-level social groups $g_{t_1}(), g_{t_2}(), \cdots, g_{t_m}()$ is available, where $t_1 > t_2 > \cdots, t_m$.

```
if v_d ∈ R then
    Forward M to v_d
else
    k = 1
    while k < m do
        if there exists v_j ∈ R within the same group at level
        k with v_d, i.e., ∃v_j ∈ R ∩ g_{t_k}(v_d) then
            Let v_k be the node in R ∩ g_{t_k}(v_d) which has con-
            tacted v_d most recently
            Forward M to v_k and k = m + 1
        else
            k = k + 1
        end if
    end while
end if
```

3.3 Routing with multi-level social groups

In Group, we only use one level of social group information to make forwarding decision. However, the choice of threshold t could affect the routing performance significantly. If t is too large, the constructed group could be too small and none relay nodes could be found in Group. If t is too small, the constructed group will include everyone and Group will regress to *FRESH*. Therefore, it makes great sense to take the advantages of wide spectrum of social relationships by considering multi-level social group information into our group-based forwarding method. We call this version of our method mGroup.

In mGroup, we consider m-level social groups $g_{t_1}(), g_{t_2}(), \ldots, g_{t_m}()$ formed by different thresholds. We assume that $t_1 > t_2 > \ldots, t_m$, thus the first level group (we call it top level) requires the strongest social tie among its members, while the m level group has the weakest social ties. See Fig. 1 for an illustration of 3-level social groups. During a round of encounter, mGroup starts with the top-level group with threshold t_1. If no node locates in the same group with the destination, mGroup will check with the second level group of the destination. This procedure continues until either it finds a relay node within the same group of the destination or m-level groups are all explored. By taking the full advantages of social groups at all levels, mGroup intends to achieve better performance than Group. Algorithm 2 shows the details of mGroup. The time complexity of Algorithm 2 at each node v_i is $O(m \cdot |R|)$ where m is the number of levels and $|R|$ is the number of nodes which v_i meets at particular time. Clearly, with a larger value of m, the time complexity increases. However, from our experiences (as shown in our simulation results reported in Sect. 4), a small value of m (such as 3 or 4) is usually sufficient. Notice that using a large number of layers with different thresholds may lead to small differences among different layers, which does not help to make smart routing decision and wastes resources in MONs.

3.4 Routing with multi-level cross-community social groups

So far we build the multi-level social groups purely from the contact graph obtained based on historical encounters, which exploit possible physical contacts between pairs of devices in the physical world. However, the proposed multi-level social group techniques can be applied to other types of social graphs and used for routing in mobile opportunistic networks, if those information are available.

For example, the data set of Infocom 2006 [17] includes answers from each participant to a questionnaire with a number of social information about this person, such as nationality, affiliation, and speaking language. These information reflects certain level of social features or relationships among users in a virtual social world. In [32], Wu and Wang showed that in this data set, the total contact times and contact durations between two individuals reduce when the social feature difference between them increases. The individuals with only one different social feature have about 36.5 % more contact times and 32.6 % longer contact durations than the individuals with two different features. Similarly, in Mei et al. [33], also found that individuals with similar social features tend to contact more often in MONs. Therefore, it is also possible to consider the social features or relationships among mobile users to improve the performance of MON routing. Both [33] and [32] directly use social features as routing metrics; however, we now consider how to use social features to discover social groups so that our social group-based routing can be applied.

Recall that our proposed method uses different values of contact strength threshold t to form multi-level social groups based on contact graph (as shown in Fig. 2a). The same approach can be easily adopted to form social groups using the social features. As illustrated in Fig. 2b, we can define the social feature strength between any two nodes as the number of common identical social features. For example, if nodes v_i and v_j only share the same nationality and affiliation, we give their social strength weight of 2. Large social strength implicitly implies strong social tie. By defining social strength among nodes, we can have a weighted social feature graph G'. Using different values of threshold t', we can then define multi-level social groups among mobile users. Since people come in contact with each other more frequently if they have more social features in common, we can prefer the nodes in the same social group with the destination as possible relay nodes. If multiple social groups coexist, the one with strongest social strength (highest in the multi-level structure) can be picked.

One advantage of using social features for routing guidance is that social feature-based routing does not need to collect and maintain routing state information. Social features are static internal features of each mobile node and usually can be obtained before the deployment of the network or during user registration phase. However, routing solely based on these static social features may not lead to great performance (as shown by our simulation results in Sect. 4.5).

Among multiple social features, some are more important than others for either routing or group formation purposes. One way to measure the importance of a social feature is to calculate its entropy value over the data set, as did in [32]. In information theory, Shannon entropy is a measure of the uncertainty associated with a random variable. We use entropy to quantify the value of the information contained in the social features. Specifically, there are n mobile users in the network, and each user has N social features, denoted as $f_1, f_2, \ldots, f_N$. For a social feature f_j with M_j possible outcomes, $\{x_i : i = 1, \ldots M_j\}$, its Shannon entropy, denoted by $E(f_j)$ is defined as

$$E(f_j) = -\sum_{i=1}^{M_j} p(x_i) log_2 p(x_i), \quad j = 1, \ldots, N$$

where $p(x_i)$ is the probability of outcome x_i in the data set. Table 1 shows the entropy of each social feature listed in the descending order for Infocom 2006 data set. It is clear that a social feature with larger entropy means better

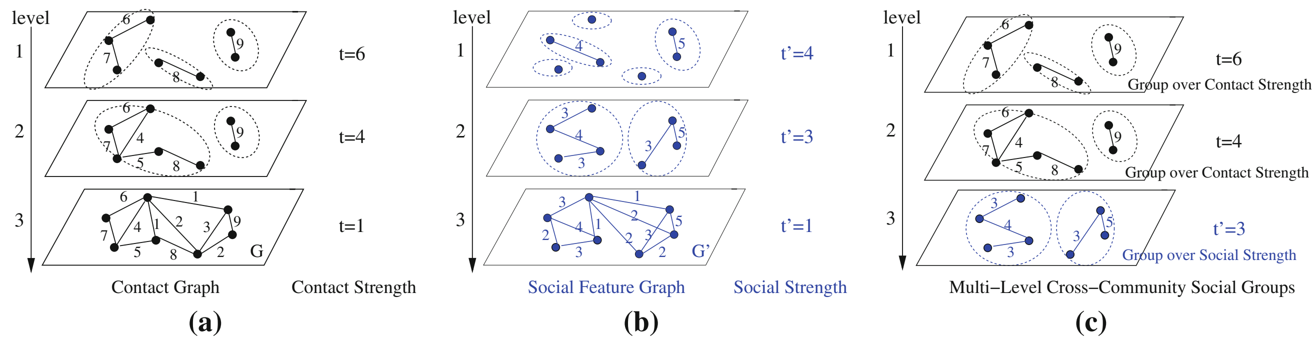

Fig. 2 Multi-level social groups based on different social graphs for the same network: **a** multi-level social groups based on the contact graph, **b** multi-level social groups based on social features, **c** multi-level cross-community social groups which include groups from heterogeneous sources

Table 1 Entropy of social features for Infocom 2006 data set

Social feature	Entropy
Graduated school	5.223
Topics	4.864
Affiliation	4.618
City	4.379
Nationality	4.333
Country	3.928
Languages	3.524
Position	1.390

distinction among mobile users. But if we use common values of social features to form social groups (*i.e.*, two users are in the same group if they share the same values of certain social features), using only social features with largest entropy may lead to small or isolate groups. This is not good for group-based routing. On the other hand, using social features with lowest entropy may lead to a huge social group, since everyone has the same values. This is also useless for routing purpose. In our simulations (Sect. 4.5), we will study the trade-off among different social features.

Notice that the social groups based on social features are distinct from the social groups based on historical encounters. The first ones are in the virtual world irrespective of physical distance among mobile users, while the second ones are in the physical world which depend on the physical proximity of users. However, both types of social groups are complementary. One of the advantages using social features is that they can be obtained before the deployment of MONs, and there is no need to maintain any states during the routing except the static social graph. On the other hand, the encounter-based social groups reflect the dynamic way people exchange information through direct, face-to-face contacts. Therefore, both types of social groups can be combined and used as a hybrid cross-community multi-level graph, as shown in Fig. 2c. This approach actually makes more sense for MONs, since modern people are living in a cross-space and multi-community coexisted world. Our simulation results in Sect. 4.5 will verify this conclusion. Notice that in Fig. 2c, we simply put the level from social features under the levels from contact graphs. This is mainly due to that in our simulations routing over social groups from contact graphs can achieve better performance than the one over social groups from social features. Thus, at current node v_i, mGroup will first explore the neighbors within the social groups from contact graphs, then try the neighbors within the groups from social features if necessary. This simple way to combine two types of social groups works well in our simulation with Infocom 2006 data set [17]. However, for other more complex data sets or cross-community scenarios, new advanced integration techniques may be needed. We leave such study as one of our future works.

4 Simulations

In this section, we conduct extensive simulations with three different realistic contact traces [15–17], which are publicly available at Crawdad [34], to evaluate our proposed method and compare it with existing opportunity-based routing schemes.

4.1 Compared routing methods

We implement our algorithm mGroup and compare it with five other existing routing methods which are listed below.

- *Epidemic* [3]: during any encounter, the message is forwarded to all encountered nodes.
- *Spray and Wait* [5]: when node v_i has $k > 1$ message replicas and meets node v_j, it gives v_j half of its replicas and keeps the other half. Initially, the source has x copies of the message. The default value of x is 10 in our simulations. If there are multiple nodes during the

encounters, v_i randomly picks one of them to share its copies.

- *FRESH* [8]: the message is only forwarded from node v_i to node v_j if v_j has met the destination more recently than v_i does. If there are multiple nodes satisfying such a condition during the encounters, v_i forwards the message to the one who has met the destination most recently.
- *Destination Frequency* [6]: the message is only forwarded from v_i to v_j if v_j has met the destination more often than v_i does. If there are multiple nodes satisfying such a condition during the encounters, v_i forwards the message to the one who has met the destination most often.
- *Greedy-Total* [9]: the message is only forwarded from v_i to v_j if v_j has more total contacts with all other nodes than v_i does. If there are multiple nodes satisfying such a condition during the encounters, v_i forwards the message to the one who has most contacts.

4.2 Routing metrics

In all experiments, we compare each algorithm using the following routing metrics.

- *Delivery ratio*: the average percentage of successfully delivered messages from the sources to the destinations.
- *Hop count*: the average number of hops during each successful delivery from the sources to the destinations.
- *Delay*: the average time duration of successfully delivered messages from the sources to the destinations.
- *Number of forwarding*: the average number of messages forwarding in the network during the whole period.

4.3 Simulation results on NUS contact trace data

In order to test our proposed forwarding method in realistic mobile opportunistic networks, we first use the NUS student contact trace [15], which was collected during the Sprint semester of 2005/2006 in National University of Singapore. There are total 22,341 students who enroll in 4,885 sessions and last for 77 session hours in this data set.

As different sessions have various start and end time and may last more than one hour, we split all sessions into unit time slot size (1 hour). Therefore, there will be total 77 time slots. If two students share the same session at a particular time slot, we consider they have contact in that time slot. Since the total number of students is too huge in the data set, we use the same method used in [10] to select a subset of students for our simulations. Contacts related to the non-selected students are ignored. Notice that if students are selected randomly, the network formed by selected students becomes too sparse for packet delivery. On the other hand, if students are selected by maximizing their connection levels (the number of shared sessions), the network becomes almost fully connected. To prevent these extremes, the selection method by [10] works as follows. The first student is randomly selected. If we already have k students, we randomly split them into two groups V_1 and V_2. Then we select the next student as the one who has highest connection to students in group V_1 and the lowest connection to students in group V_2 among the students that are not yet selected. Here, the level of connection is the average number of shared sessions. We put the selected student in group V_1. This procedure is repeated (new group V_1 and group V_2 will be generated from $k+1$ selected nodes) until we have the required number of students. A clustering factor C is defined as $\frac{|V_1|}{|V_1|+|V_2|}$, which is used to control the degree of connectivity in the network. In our simulations, we randomly choose C from 0.1 to 0.9 as [10] did. For more details about the selection method, please refer to [10].

Table 2 Parameters used for NUS data set

Parameter	Value or range
Total number of students/sessions	22,341/4,885
Total number of time slots	77
Number of selected students	100–600
Clustering factor C	0.1–0.9
Number of routing tasks	50 or 100
Number of message replicas allowed	1, 10 or 20
Number of levels of social graph	1 or 4

For routing tasks, we randomly choose 50 or 100 pairs of selected students as the sources and destinations. All results are reported as the average of these 50 or 100 tasks. We test both single-copy and multiple-copy routing versions of all methods. For the multi-copy case, we allow the number of replicas at the source to be either 10 or 20. Table 2 summarizes the parameters used in our experiments.

In total, we have conducted four sets of simulations. In all simulations, we increase the number of selected students from 100 to 600. Fifty routing tasks are performed in each setting of the first three sets of simulations, while 100 routing tasks are used in the last set of simulations. For all methods in the first three sets of simulations, we use the first 40 sessions data as historical data to obtain statistics (such as social groups or encounter counts), and routing tasks are performed for the last 37 sessions to evaluate the routing performances. For those in the last set, we use a sliding window with size of 30 sessions to perform the

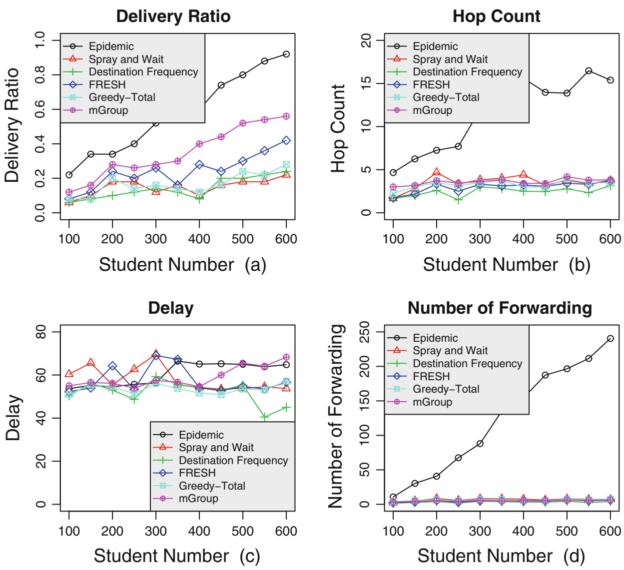

Fig. 3 Simulation results of all different routing algorithms

group formation and routing methods are evaluated over all sessions.

In the first set of simulations, we consider all methods (using their multi-copy versions in which the total number of copies is limited to 10, except for epidemic routing). Figure 3 shows all results. It is clear that the delivery ratio is increasing as the number of students (devices) increases. This is reasonable since denser networks provide more opportunities for message delivery. From Fig. 3a, we can see that our proposed mGroup algorithm achieves better delivery ratio than any others except for epidemic routing. Notice that even though epidemic routing has the best delivery ratio, it costs extremely large amount of forwarding as shown in Fig. 3d. In terms of hop count, delay, and number of forwarding, mGroup is at the similar level with those of other opportunity-based methods.

In the second set of simulations, we compare the performance of our group-based methods using either a single-level social group (Group) or multi-level social groups (mGroup). For Group, we use different threshold values $t = 5, 10, 15, 20$. For mGroup, we use all these 4-level of social groups. Figure 4 shows the results in which Group_t denotes Group method with threshold value t. For Group_t with single-level social group, larger threshold t leads to higher delivery ratio since it provides the information of stronger social ties. However, overall mGroup with multi-level social groups has the highest delivery ratio. This confirms our original conjecture of better performance with more information.

In the third set of simulations, we consider both single-copy and multi-copy versions of our methods. We allow different numbers of copies (1, 10, or 20) during the routing. Figure 5 shows the comparison in which mGroup_x denotes mGroup method with limitation of x numbers of copies. It is obvious that with more message copies, all methods can achieve higher delivery ratio but increase the number of forwarding too. There is clearly a trade-off between number of copies and forwarding overhead.

In the last set of simulations, instead of using static historical encounters to perform the social grouping, a sliding window approach is used. For current time slot, the grouping is done with the encounter information in the previous w time slots. In other words, the social groups are

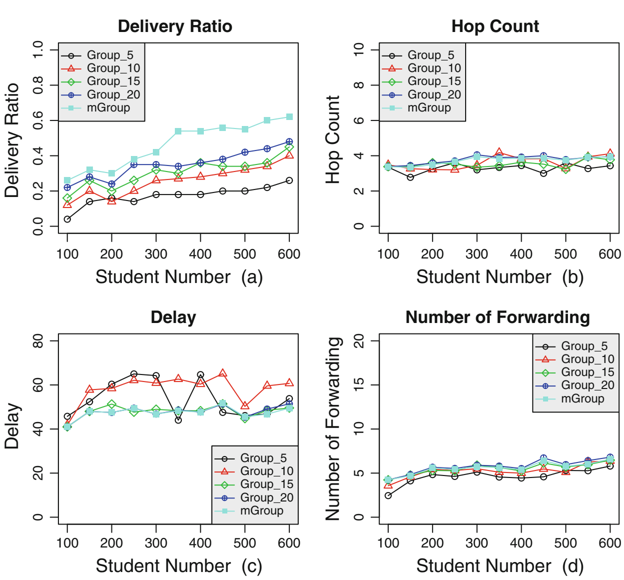

Fig. 4 Simulation results of mGroup and Group routing

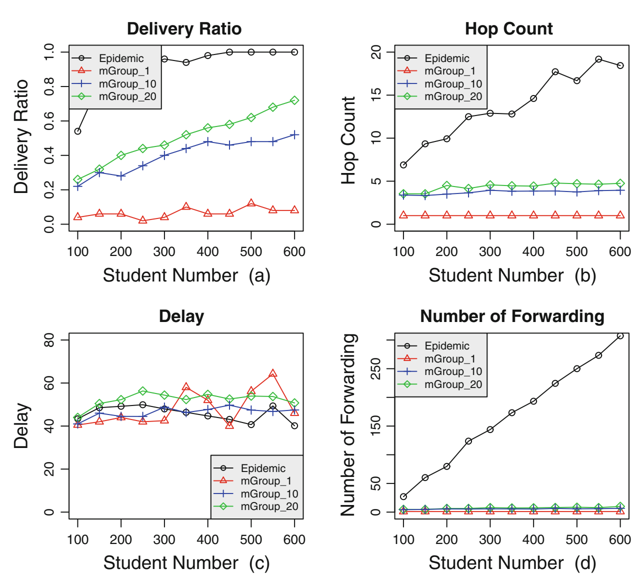

Fig. 5 Simulation results of mGroup routing with different numbers of copies

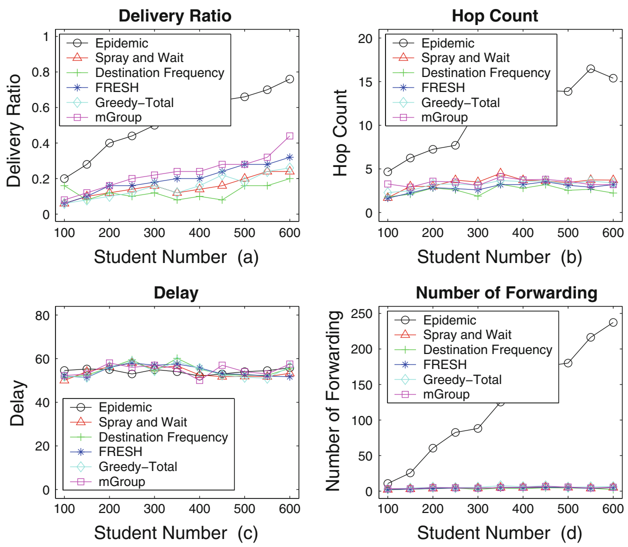

Fig. 6 Simulation results of all different routing algorithms with sliding window (window size $w = 30$)

dynamic and time evolving. We let $w = 30$ and other settings be the same with those in the first set of simulations except for 100 random routing tasks. Figure 6 gives the average performances of different routing algorithms over 100 routing tasks. The results are consistent with those using static historical encounter information. This demonstrates that the proposed method can be used in online fashion. We also test different window sizes, the results are similar thus ignored here.

4.4 Simulation results on Cambridge Bluetooth trace data

We also test the performances of proposed single-level social group (Group) and multi-level social groups (mGroup) protocols using a real-life Bluetooth trace data set. The Bluetooth trace data [16] were collected from a group of mobile users (i.e., students from Cambridge University) who were asked to carry iMotes with them for two months starting from October 28, 2005, in addition with a number of stationary nodes in various locations that were expected many people to visit, such as grocery stores, pubs, market places, and shopping centers in and around the city of Cambridge, UK. The Bluetooth trace data consist of measurement data from 36 mobile participants and 18 fixed locations. In order to discover the social relationships among mobile users, we only use tracing contacts between 36 mobile students in this set of simulations.

We split the contact time duration into unit time slot size (one hour) and test the performances of proposed single-level social group (Group) and multi-level social groups (mGroup) protocols together with other existing routing methods mentioned in Sect. 4. A in randomly chosen 120 time slots. For all routing protocols, we use the first 40 hours data as historical data to obtain social group information and evaluate the performances of routing tasks for the remaining 80 hours. In the simulation, each student tries to send message to other 35 students, and thus, there are $36 \times 35 = 1260$ source and destination pairs in total. The simulation results are shown in Table 3 which are the average of these 1260 routing tasks. The number of message replicas allowed is set to 10 except for epidemic routing. The simulation results for Bluetooth trace data are consistent with those for NUS trace data, such as for Group_t with single-level social group method, larger threshold t leads to higher delivery ratio; 3-level social groups mGroup protocol have the highest delivery ratio among all other routing methods except for epidemic routing, but the number of message forwarding of mGroup is much less than epidemic routing and mGroup is at the similar level in terms of hop count, delay, and number of forwarding with those of other opportunity-based methods.

4.5 Simulation results on Infocom 2006 Bluetooth trace data with social features

Finally, to test the cross-community social group idea in mGroup, we use the Infocom 2006 trace data [17]. This data set includes Bluetooth sightings by groups of users (*i.e.*, 79 participants) carrying iMotes for four days during Infocom 2006 conference in Barcelona, Spain. In addition to the Bluetooth contact information among participants, it also includes social features of each participant, which are the statistics of participants' information returned from a questionnaire form. Since some social features in participants' questionnaire forms are blank, we extract eight social features from the original data set: *nationality, graduated school, languages, current affiliation, current position, city of residence, country of residence*, and *interested topics*, whose entropy values are summarized in Table 1.

There are 74,981 contacts between 79 participants over a period of 337,418 s. We divide the period using time slot with length of one hour and test the routing performance over randomly chosen 120 time slots. Again, we use the first 40 hours data as historical data to obtain social group information and then evaluate the performance of routing tasks over the remaining 80 hours. In the simulations, each participant tries to send a message to all other 78 participants. The number of message replicas allowed is set to 10 except for epidemic routing.

Table 4 shows the performances of existing methods and Group/mGroup over social groups generated from contact graphs. Here, Group_t is the method over a single-level social group with the threshold value of encounter t,

Table 3 Simulation results on Cambridge Bluetooth data set

	Epidemic	Spray and wait	Destination frequency	FRESH	Greedy-total
Delivery ratio	0.68	0.58	0.61	0.60	0.58
Hop count	1.9	3.1	2.2	2.7	3.4
Delay	68	69	69	69	69
Number of forwarding	12.6	5.6	2.9	4.4	5.0
		Group_5	Group_10	Group_15	mGroup
Delivery ratio		0.60	0.63	0.66	0.67
Hop count		2.9	2.9	3.1	3.1
Delay		69	68	69	51
Number of forwarding		4.0	4.0	4.4	4.2

Table 4 Simulation results on Infocom 2006 data set (social groups from contact graphs)

	Epidemic	Spray and wait	Destination frequency	FRESH	Greedy-total
Delivery ratio	0.70	0.38	0.25	0.46	0.39
Hop count	3.8	4.0	1.6	3.6	3.8
Delay	54	57	56	58	56
Number of forwarding	14.6	6.9	2.0	6.0	3.6
		Group_10	Group_50	Group_100	mGroup
Delivery ratio		0.37	0.41	0.44	0.45
Hop count		3.7	3.8	3.8	3.8
Delay		56	56	59	60
Number of forwarding		6.9	7.5	6.9	6.9

Table 5 Simulation results on Infocom 2006 data set (social groups from social features)

	mGroup	Group'_2(t)	Group'_2(m)	Group'_2	Group'_4	Group'_6	mGroup'
Delivery ratio	0.45	0.23	0.27	0.29	0.32	0.35	0.39
Hop count	3.6	3.8	3.8	3.5	3.3	3.6	3.8
Delay	60	56	60	58	59	61	63
Number of forwarding	6.9	6.3	6.0	6.8	6.3	5.9	6.0

Table 6 Simulation results on Infocom 2006 data set (social groups from both contact graphs and social features)

	mGroup	mGroup+Group'_2	mGroup+Group'_4	mGroup+Group'_6	mGroup+mGroup'
Delivery ratio	0.45	0.46	0.48	0.50	0.55
Hop count	3.8	3.8	3.9	3.6	3.8
Delay	60	66	62	62	62
Number of forwarding	6.9	6.0	5.3	6.2	5.9

and mGroup is the method over 3-level social groups with threshold values of encounters $t = 100, 50, 10$ from the top level to the lower level. Clearly, mGroup achieves better performance than Group_t.

We then implement both Group and mGroup over social groups generated purely from static social features in the data set. As we described in Sect. 3, we measure the number of common social features among nodes. We use Group'_t' to represent Group method with threshold value t' over the social feature strength. mGroup' is mGroup method over 3-level social groups with threshold values of social feature strength $t' = 6, 4, 2$ from the top level to the lower level. For Group'_2 ($t' = 2$), we have three different variations: Group'_2 (all eight social features listed in Table 1 are considered), Group'_2(t) (only the top four social features listed in Table 1 are considered), and Group'_2(m) (only the middle four social features listed in Table 1 are considered). In other words, Group'_2, Group'_2(t), and Group'_2(m) are three one-level social group routing methods where during the group formation phase, two participants will be put into the same group if they have at least two common identical social features among all eight social features or the top four social features or the middle four social features listed in Table 1, respectively. Table 5 shows all simulation results. First, clearly routing purely over social groups from static social features (Group'_t' or mGroup') achieves lower delivery ratio than the method over social groups from contact graphs mGroup. This is reasonable since the contact graphs generated based on Bluetooth traces are more directly reflecting the physical contact opportunities used for packet delivery. Second, for single-level version Group'_t', larger threshold value t' leads to higher delivery ratio since it provides the information of stronger social ties. Third, mGroup' can achieve better performance than single-level version Group'_t' with more multi-level information. Forth, Group'_2 outperforms Group'_2(t) and Group'_2(m). Thus, considering more social features is helpful. Last, compared with Group'_2(t), Group'_2(m) has better performances though they both consider four social features. This shows that it does not necessarily lead to better performance using social features with higher entropy. Notice that if the entropy of a social feature is very large, it is hard to find common values of that social feature among users.

Finally, we implement new hybrid versions of mGroup by adding additional levels of social groups from social features under the social groups from contact graphs. Here, mGroup+Group'_2 represents a hybrid 4-level cross-community social groups, which combines the 3-level social groups mGroup (with threshold values of encounters $t = 100, 50, 10$) and an additional level Group'_2 of social groups (with 2 common social features) as the bottom level. Similarly, we have the other two hybrid 4-level cross-community social groups: mGroup+Group'_4 and mGroup+Group'_6. We also test mGroup+mGroup' which is a hybrid 6-level cross-community social groups with 3-level from mGroup and the other 3-level from mGroup' (3-level social groups with threshold values of social feature strength $t' = 6, 4, 2$). Table 6 shows the simulation results. It is clear that with new additional information from social features, our multi-level social group method can further improve the delivery ratio while maintain similar level of other metrics. This confirms the complementary properties of two types of social groups obtained from both physical and virtual worlds and the benefits of combining useful social information cross-space/community.

5 Conclusion

Mobile opportunistic networking is a new emerging communication system which takes the advantages of any possible contact opportunities to deliver data among mobile devices. Routing in such networks is a challenging problem. In this paper, we propose a new group-based routing method which forwards message based on multi-level cross-community social group information. Our simulation results demonstrate the great performance of the proposed method and the advantages of considering diverse social relationships among nodes during relay selection. In this paper, we only use a few simple and static social features available in the data set (Infocom 2006 trace data [17]) to test our cross-community methods. We leave exploring more complex social group analysis using more enriched social features (such as Facebook friendship or location profiles) to achieve further performance improvement as one of our future works, when such information becomes available in public data sets.

Acknowledgments The work of F. Li is partially supported by the National Natural Science Foundation of China under Grant 60903151, Beijing Natural Science Foundation under Grant 4122070 and Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry. The work of Y. Wang is supported in part by the US National Science Foundation under Grant No. CNS-0915331 and CNS-1050398.

References

1. Hui P, Chaintreau A, Scott J, Gass R, Crowcroft J, Diot C (2005) Pocket switched networks and human mobility in conference environments. In: Proceedings of ACM SIGCOMM workshop on DTN (WDTN)
2. Guo B, Zhang D, Yu Z, Zhou X, Zhou Z (2012) Enhancing spontaneous interaction in opportunistic mobile social networks. Commun Mob Comput 1:6
3. Vahdat A, Becker D (2000) Epidemic routing for partially connected adhoc networks. Technical report. CS-200006, Duke University
4. Hui P, Crowcroft J, Yonek E (2008) Bubble rap: social-based forwarding in delay tolerant networks. In: Proceedings of ACM MobiHoc
5. Spyropoulos KT, Psounis, Raghavendra C (2005) Spray and wait: an efficient routing scheme for intermittently connected mobile networks. In: Proceedings of ACM SIGCOMM workshop on DTN (WDTN)
6. Erranmilli V, Crovella M, Chaintreau A, Diot C (2008) Delegation forwarding. In: Proceedings of ACM MobiHoc
7. Daly EM, Haahr M (2007) Social network analysis for routing in disconnected delay-tolerant MANETs. In: Proceedings of ACM MobiHoc
8. Dubois-Ferriere H, Grossglauser M, Vetterli M (2003) Age matters: efficient route discovery in mobile ad hoc networks using encounter ages. In: Proceedings of ACM MobiHoc
9. Erramilli V, Crovella M (2007) Diversity of forwarding paths in pocket switched networks. In: Proceedings of ACM IMC
10. Liu C, Wu J (2009) An optimal probabilistic forwarding protocol in delay tolerant networks. In: Proceedings of ACM MobiHoc
11. Hui P, Crowcroft J (2007) How small labels create big improvements. In: Proceedings of international workshop on intermittently connected mobile Ad hoc networks (ICMAN)

12. Bulut E, Szymanski BK (2010) Friendship based routing in delay tolerant mobile social networks. In: Proceedings of IEEE GLOBECOM
13. Zhu Y, Xu B, Shi X, Wang Y (2013) A survey of social-based routing in delay tolerant networks: positive and negative social effects. IEEE Commun Surv Tutor 15(1):387–401
14. Guo B, Zhang D, Yu Z, Zhou X (2012) Hybrid SN: interlinking opportunistic and online communities to augment information dissemination. In: Proceedings of 9th IEEE international conference on ubiquitous intelligence and computing (UIC'12)
15. Srinivasan V, Motani M, Ooi WT (2006) Analysis and implications of student contact patterns derived from campus schedules. In: Proceedings of ACM MobiCom
16. Leguay J, Lindgren A, Scott J, Friedman T, Crowcroft J (2006) Opportunistic content distribution in an urban setting. In: Proceedings of ACM SIGCOMM workshop on challenged networks (CHANTS)
17. Scott J, Gass R, Crowcroft J, Hui P, Diot C, Chaintreau A (May 2009) CRAWDAD trace cambridge/haggle/imote/infocom2006 (v.2009-05-29), Downloaded from http://crawdad.cs.dartmouth.edu/cambridge/haggle/imote/infocom2006
18. Zhao L, Li F, Zhang C, Wang Y (2012) Routing with multi-level social groups in mobile opportunistic networks. In: Proceedings of IEEE global telecommunications conference (GlobeCom 2012)
19. Li F, Zhang C, Gao Z, Zhao L, Wang Y (2012) Social feature enhanced group-based routing for wireless delay tolerant networks. In: Proceedings of the 8th international conference on mobile ad-hoc and sensor networks (MSN 2012)
20. Zhang Z (2006) Routing in intermittently connected mobile ad hoc networks and delay tolerant networks: overview and chalenges. IEEE Commun Surv Tutor (1):24–37
21. Leguay J, Friedman T, Conan V (2005) DTN routing in a mobility pattern space. In: Proceedings of ACM SIGCOMM workshop on DTN
22. Hillery GA (1955) Definitions of community: areas of agreement. Rur Sociol 20:111
23. McMillan DW, Chavis DM (1986) Sense of community: a definition and theory. J Commun Psychol 14:6–23
24. Perkins DD, Florin P, Rich RC, Wandersman A, Chavis DM (1990) Participation and the social and physical environment of residential blocks: crime and community context. Am J Commun Psychol 18:83–115
25. Okasha S (2005) Altruism, group selection and correlated interaction. Br J Philos Sci 56:703–725
26. Newman MEJ (2006) Modularity and community structure in networks. Proc Nat Acad Sci 103:8577–8578
27. Hui P, Yoneki E, Chan SY, Crowcroft J (2007) Distributed community detection in delay tolerant networks. In: Proceedings of the ACM international workshop on mobility in the evolving internet architecture (MobiArch)
28. Yoneki E, Hui P, Crowcroft J (2007) Visualizing community detection in opportunistic networks. In: Proceedings of ACM MobiCom workshop on challenged networks (CHANTS)
29. Hui P, Yoneki E, Crowcroft J (2009) Identifying social communities in complex communications for network efficiency. In: Proceedings of the international conference on complex sciences: theory and applications (Complex)
30. Chan SY, Hui P, Xu K (2009) Community detection of time-varying mobile social networks. In: Proceedings of the international conference on complex sciences: theory and applications (Complex)
31. Guo B, He H, Yu Z, Zhang D, Zhou X (2012) GroupMe: Supporting group formation with mobile sensing and social graph mining. In: Proceedings of the 9th international conference on mobile and ubiquitous systems: computing, networking and services (MobiQuitous'12)
32. Wu J, Wang Y (2012) Social feature-based multi-path routing in delay tolerant networks. In: Proceedings of the 31st IEEE international conference on computer communications (IEEE INFOCOM)
33. Mei A, Morabito G, Santi P, Stefa J (2011) Social-aware stateless forwarding in pocket switched networks. In: Proceedings of the 30th IEEE international conference on computer communications (IEEE INFOCOM)
34. CRAWDAD data set, http://crawdad.cs.dartmouth.edu/